



Ecole d'ingénieurs et d'architectes de Fribourg
Hochschule für Technik und Architektur Freiburg

DIPLOMA ABSTRACT | FRIBOURG NOVEMBER 08

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES

COMPUTER SCIENCE

Virtual Machines Converter for a Virtual Machine Logbook in ATLAS

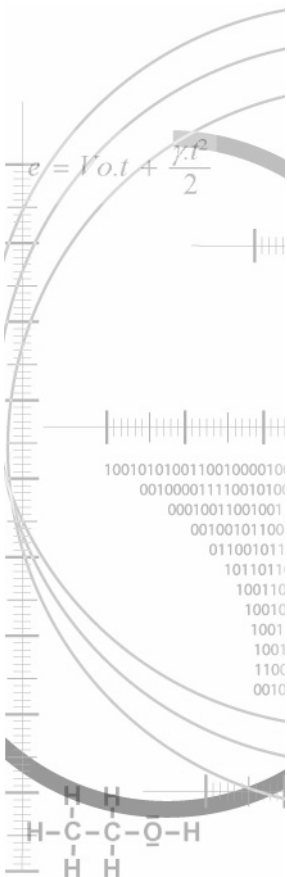
ACRONYM	VM Atlas 1
APPLICANT	Lawrence Berkeley National Laboratory, USA
STUDENT	Andrea Cavalli
PROFESSORS	Ottar Johnsen, Frédéric Bapst
EXPERT	Peter Kropf
N°	D08I04
TYPE	Diploma project
CONTACT	P. Calafiura

ATLAS is both one of the detector in the LHC particle accelerator at CERN, and the related software (organized in a thousand packages) used by physicists when writing data analysis programs. Sharing the work among the large ATLAS community can be hard, because the different working environments can lead to different behaviors. In this context, virtualization appears as a promising technology, enabling the exchange of the full context of a particular program run. But this raises new questions around the management of numerous virtual machines among the ATLAS community.

This leads to the idea of the Virtual Machine Logbook (VML), an application whose goal is to simplify the share of the work environments used in the ATLAS experiment. The logbook will be a repository which stores different environments used by physicists. VML entries represent virtual machines that can be added, checked out, or removed, thus helping managing the history of the work, and the sharing between users. Two EIA-FR students will develop VML at Berkeley, during a 30-weeks period.

This particular project deals with converting the virtual machines between various representations in VML; among others, it covers the following aspects:

- "virtual machine" to "VML entry"
- "VML entry" to "virtual machine"
- generic conversions
- ATLAS-specific conversions



Introduction

The goal of this project is to develop Virtual Machine Logbook (VML) in collaboration with Julien Poffet, another student of the EIA-FR. The development of the application is done at the Lawrence Berkeley Laboratory in California.

VML is a versioning system conceived for the physicists involved in the ATLAS experiment of the LHC particle accelerator at CERN. It allows keeping a history of the work done by the physicists with virtual machines.

VML is still in development. This abstract explains the work done so far and the goals to achieve by the end of this project in February 2009.

Context

This project deals with virtualization, which is the technique used to "simulate" a real computer inside an application. A computer which is virtualized is called virtual machine (VM); inside a VM runs an operating system, then programs.

At CERN, physicists are starting to develop their analysis programs inside virtual machines instead of on real computers, because this will simplify the sharing of their work.

Virtual Machine Logbook

VML provides a system to keep the history of the work done by the physicist inside a virtual machine. Virtual machines can be saved as VML entries into a repository; then VML can restore the state of an old version of the VM using the corresponding entry.

VML entries

All files created inside a virtual machine are saved on its virtual disk. On the real computer where the VM runs, this disk is a file – a big file. Saving a virtual machine essentially means to save its virtual disk. Since virtual disks are very big files, VML provides a way to reduce the size of the entries stored into the repository.

This project focuses on how to convert a virtual machine into a space-saving VML entry and how to restore a VM from a VML entry.

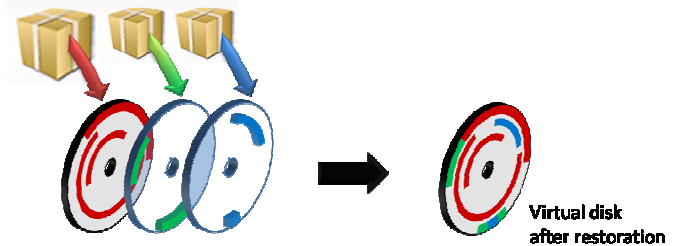


Incremental Backups

A generic way to convert a virtual machine in a VML entry is to use incremental backups. With

this technique, VML saves into the entry only the changes done on the virtual disk since the last version. Only the first version contains the whole virtual disk and the following entries are considerably smaller.

The restoration of an entry is done by applying to the virtual disk all the changes contained in each entry.



Environment Backups

With the environment backups, only important data of the virtual disk is saved into an entry. Important data are the files modified by the physicist and the configurations of the applications he uses to develop his applications. This data is the environment in which the physicist works.

To restore an old version of the virtual machine, VML replaces the current environment of the VM with the old one by copying back the contents of the entry to the virtual machine.

Environment backups will be implemented during the next months of the project. VML will provide this for VMs specific to the ATLAS experiment.

Entries Sharing

VML entries are simply shareable among VML users. Since the physicists of the CERN work on the same kind of virtual machine, the most part of the virtual disk they use is common. When the physicist shares his virtual machine using VML entries, he only shares the modifications he did on the virtual disk.

Conclusion

With the conversion of the virtual machines in VML entries, versioning and sharing virtual machines is efficient in terms of data size.

The first part of this project focused on the communication with the virtual machines and also on the management and the storage of the entries.

The planning for the next months is to make VML more efficient with the VMs used by the physicists (by implementing the environment backups), more user-friendly and capable to manage more virtual machines formats.